IN THE UNITED STATES DISTRICT COURT FOR THE SOUTHERN DISTRICT OF WEST VIRGINIA

CHARLESTON DIVISION

OHIO VALLEY ENVIRONMENTAL COALITION,
WEST VIRGINIA HIGHLANDS CONSERVANCY,
WEST VIRGINIA RIVERS COALITION,
and SIERRA CLUB,

Plaintiffs,

v. CIVIL ACTION NO. 2:15-1371

FOAL COAL COMPANY, LLC,

Defendant.

MEMORANDUM OPINION AND ORDER

On March 13 and 14, 2017 the Court held a bench trial in this case on liability issues,[1] and the parties timely conducted post-trial briefing. As explained more fully in this Memorandum Opinion, the Court **FINDS** that Plaintiffs have established, by a preponderance of the evidence, that Defendant has violated its permits by discharging high levels of ionic pollution, as measured by conductivity, into Shanty Branch and Elick Hollow, which have caused or materially contributed to a significant adverse impact to the chemical and biological components of the applicable streams' aquatic ecosystems, in violation of the narrative water quality standards that are incorporated into those permits. The Court also **FINDS** Plaintiffs have not established, by a preponderance of the evidence, that Defendant violated its permit as it relates to ionic pollution flowing into Leatherwood Creek. Both Shanty Branch and Elick Hollow flow into Leatherwood

[1] Pursuant to the Scheduling Order in this case, the case is proceeding in two phases: Phase I will resolve issues of jurisdiction and liability; and Phase II, if necessary, will address the remedy. ECF No. 9.

Creek, as do a number of other streams carrying mine drainage. Plaintiffs have not quantified pollution contributions from Fola to Leatherwood Creek by way of Shanty Branch and Elick Hollow and therefore have not met their burden on this issue.

I. Background

A. Regulatory Framework

Plaintiffs bring this action pursuant to the citizen suit provisions of the Federal Water Pollution Control Act ("Clean Water Act" or "CWA") and the Surface Mining Control and Reclamation Act ("SMCRA"). Plaintiffs allege that Defendant Fola Coal Company, LLC ("Fola") violated these statutes by discharging excessive amounts of ionic pollution, measured as conductivity, into the waters of West Virginia in violation of their National Pollution Discharge Elimination System ("NPDES") permits and their West Virginia Surface Mining Permits.

The primary goal of the CWA is "to restore and maintain the chemical, physical, and biological integrity of the Nation's waters." 33 U.S.C. § 1251(a). To further this goal, the Act prohibits the "discharge of any pollutant by any person" unless a statutory exception applies; the primary exception is the procurement of an NPDES permit. 33 U.S.C. §§ 1311(a), 1342. Under the NPDES, the Environmental Protection Agency ("EPA") or an authorized state agency can issue a permit for the discharge of any pollutant, provided that the discharge complies with the conditions of the CWA. 33 U.S.C. § 1342. A state may receive approval to administer a state-run NPDES program under the authority of 33 U.S.C. § 1342(b). West Virginia received such approval, and its NPDES program is administered through the West Virginia Department of Environmental Protection ("WVDEP"). 47 Fed. Reg. 22363-01 (May 24, 1982). All West Virginia NPDES permits incorporate by reference West Virginia Code of State Rule § 47-30-5.1.f, which states that "discharges covered by a WV/NPDES permit are to be of such quality so as not to cause

violation of applicable water quality standards promulgated by [West Virginia Code of State Rules § 47-2].” This is an enforceable permit condition. *Ohio Valley Envtl. Coal. v. Fola Coal Company, LLC*, 845 F.3d 133, 142 (4th Cir. 2017) (*Fola*).

Coal mines are also subject to regulation under the SMCRA, which prohibits any person from engaging in or carrying out surface coal mining operations without first obtaining a permit from the Office of Surface Mining Reclamation and Enforcement (“OSMRE”) or an authorized state agency. 30 U.S.C. §§ 1211, 1256, 1257. A state may receive approval to administer a state-run surface mining permit program under the authority of 30 U.S.C. § 1253. In 1981, West Virginia received conditional approval of its state-run program, which is administered through the WVDEP pursuant to the West Virginia Surface Coal Mining and Reclamation Act (“WVSCMRA”). W. Va. Code §§ 22-3-1 to -33; 46 Fed. Reg. 5915-01 (Jan. 21, 1981). Regulations passed pursuant to the WVSCMRA require permittees to comply with the terms and conditions of their permits and all applicable performance standards. W. Va. Code R. § 38-2-3.33.c. One of these performance standards requires that mining discharges “shall not violate effluent limitations or cause a violation of applicable water quality standards.” W. Va. Code R. § 38-2-14.5.b. Another performance standard mandates that “[a]dequate facilities shall be installed, operated and maintained using the best technology currently available . . . to treat any water discharged from the permit area so that it complies with the requirements of subdivision 14.5.b of this subsection.” W. Va. Code R. § 38-2-14.5.c.

West Virginia’s water quality standards are violated if wastes discharged from a surface mining operation “cause . . . or materially contribute to” 1) “[m]aterials in concentrations which are harmful, hazardous or toxic to man, animal or aquatic life” or 2) “[a]ny other condition . . . which adversely alters the integrity of the waters of the State.” W. Va. Code R. § 47-2-3.2.e, -3.2.i.

Additionally, "no significant adverse impact to the chemical, physical, hydrologic, or biological components of aquatic ecosystems shall be allowed." W. Va. Code R. § 47-2-3.2.i.

B. Factual Background

This controversy concerns discharges from a surface mine along the southern portion of the Leatherwood Creek watershed. The mine at issue, Fola's Monoc #2 Surface Mine, is located in Clay and Nicholas Counties, West Virginia and is situated on the other side of Leatherwood Creek from Fola's Surface Mine No.2 and near Fola's Surface Mines No. 4A and No. 6, all situated along Leatherwood Creek. Stipulation, ¶ 1, Pls.' Ex. 51, ECF No. 72-24. The latter three mines were the subject of prior litigation between Plaintiffs, except the West Virginia Rivers Coalition, and Fola. *See Ohio Valley Envtl. Coal. v. Fola Coal Co., LLC*, 120 F. Supp. 3d 509 (S.D. W. Va. 2015) (*Leatherwood*). In that case, this Court found that Fola violated it CWA and SMCRA permits for the No. 2 and No. 6 mines by discharging highly conductive water into two tributaries of Leatherwood Creek. *Id.* at 544–46.

The Monoc #2 mine area contains two valley fills. Valley Fill #1 partially fills Elick Hollow, which drains into Pond #1 and then from Outlet 005 into Elick Hollow of Leatherwood Creek. Pls.' Ex. 51 ¶ 2. Valley Fill #2 partially fills Shanty Branch, which drains into Pond #2 and then from Outlet 011 into Shanty Branch of Leatherwood Creek. *Id.*

Defendant's mining activities at the Monoc #2 mine are regulated under West Virginia Surface Mining Permit S6019-89 and WV/NPDES Permit WV1009290. *Id.* ¶¶ 4, 5. The permits were transferred from Vandalia Resources to Fola in 2002 in the case of the former and 2004 for the latter. WVDEP reissued WV/NPDES Permit WV1009290 in April 2013. *Id.* ¶ 5. It limits discharges from Outlets 005 and 011. *Id.* Outlet 005 is the only point source in Elick Hollow and Outlet 011 is the only point source in Shanty Branch. *Id.*

Before mining began, Fola reported that conductivity in Elick Hollow measured 35 µS/cm, while conductivity in Shanty Branch measured 44 µS/cm. *Id.* ¶ 7. After mining began, Fola measured levels of conductivity at Outlets 005 and 011 and at instream monitoring points in Leatherwood Creek. *Id.* ¶ 8. From 1992 to 2000 Fola measured highly conductive discharges of water from Outlets 005 and 011. *Id.* (showing discharges from Outlets 005 and 011 consistently ranging from 1,000 µS/cm up to 4,000 µS/cm.). Instream levels of conductivity in Leatherwood Creek downstream from Outlets 005 and 011 were also extremely high compared to pre-mining measures. *Id.* (showing instream conductivity consistently ranging from approximately 500 µS/cm to 3,000 µS/cm with a general trend toward increasing conductivity over time). From 2008 to the present Fola has measured conductivity at Outlets 005 and 011 consistently ranging from approximately 1,000 µS/cm to approximately 2,000 µS/cm. *Id.* ¶ 9.

The most recent data collected shows that conductivity of the discharge from Outfall 011, which flows into Shanty Branch, measured 1,860 µS/cm and 2,060 µS/cm in April and October 2016, respectively. Pls.' Ex. 42, ECF No. 72-18. Instream conductivity levels in Shanty Branch measured over the same period (April and October 2016) were 1,860 µS/cm. *Id.* Conductivity at Outfall 005, which flows into Elick Hollow, measured 1,897 µS/cm and 2,030 µS/cm in April and October 2016, respectively. Pls.' Ex. 39, ECF No. 72-16. During that same period, instream conductivity in Elick Hollow measured 1,854 µS/cm and 2,060 µS/cm. *Id.*

Before mining began at the Monoc #2 mine sulfate levels were similarly quite low. *Id.* ¶ 7. In May 1988, Fola reported sulfate levels in Elick Hollow at 9.88 mg/L. *Id.* At the same time, Fola reported sulfate levels in Shanty Branch at 23.04 mg/L. *Id.* The most recent data show that sulfate levels in Elick Hollow and Shanty Branch have increased to approximately 1,100 mg/L, which is identical to the sulfate level of the discharges from Outfalls 005 and 011. Pls.' Ex. 39, ECF No.

72-16; Pls.' Ex. 42, ECF No. 72-18. The elevated levels of sulfates in both streams are consistent with data collected by Fola since 2008. *See* Stipulation ¶ 9, Pls.' Ex. 51, ECF NO. 72-24 (showing sulfate levels at Outfalls 005 and 011 between approximately 800 mg/L and 1700 mg/L).

On February 27, 2017 and memorialized in a Memorandum Opinion and Order dated April 4, 2017, the Court granted Plaintiffs' Partial Motion for Summary Judgment, finding Plaintiffs had standing to bring this suit and complied with all pre-suit requirements. Memorandum Opinion and Order, ECF No. 74. In addition, the Court found Fola was collaterally estopped from relitigating the issue of general causation of impairment by ionic pollution composed of a consistent mix of ions typical of alkaline mine drainage in the Appalachian region. *Id.*; *see also Ohio Valley Envtl. Coal., Inc. v. Fola Coal Co., LLC*, 82 F. Supp. 3d 673, 687 (S.D. W. Va. 2015) (*Stillhouse*) (finding the ions Ca^{2+}, Mg^{2+}, SO_4^{2-} and HCO_3^- as the consistent mix of ions indicative of alkaline mine drainage in Appalachia that causes impairment).

The remaining issue and the subject of the trial was whether discharges of ionic pollution impaired the streams at issue in this case.

II. Liability

To prove Fola violated its permit, Plaintiffs must show by a preponderance of the evidence that the streams into which Outfalls 005 and 011 discharge are impaired and that ionic pollution from those discharges is causing or materially contributing to that impairment. *See Stillhouse*, 82 F. Supp. 3d at 396; *Leatherwood*, 120 F. Supp. 3d at 545 (finding Fola in violation of its NPDES permits where plaintiffs showed the streams at issue were (1) impaired; (2) the discharges from Fola share the same ionic composition as that known to cause impairment in Appalachian streams; (3) conductivity levels were far in excess of thresholds identified by EPA and other scientific literature known to cause stress to aquatic animals; (4) Fola's mining operation was the only land

use that could have produced the ionic pollution; and (5) changes in the biological community show the loss of conductivity-intolerant organisms). The narrative water quality standards only require Plaintiffs to prove that Fola is "causing or materially contributing" to impairment. W. Va. Code R. § 47-2-3.2.e, -3.2.i. As this Court has found in prior litigation, "materially contribute" does not require Plaintiffs to conclusively eliminate "contributions by additional factors in a dynamic system." *Stillhouse*, 82 F. Supp. 3d at 685–86. "Plaintiffs' burden is not scientific certainty but legal sufficiency." *Id.* at 694 (quoting *Ferebee v. Chevron Chem. Co*, 736 F.2d 1529, 1536 (D.C. Cir. 1984)).

A. Determining Violations of Narrative Water Quality Standards

In its post-trial briefing, Fola contends that Plaintiffs should not be able to rely on the Genus Level Index of Most Probable Stream Status ("GLIMPSS")[2] to show impairment. Instead, Fola argues for the exclusive use of the West Virginia Stream Condition Index ("WVSCI")[3] as the sole determination of compliance with West Virginia's narrative water quality standards. To use GLIMPSS now, Fola maintains, would change the water quality standard without fair notice of the change, and would usurp West Virginia's authority to determine which biological assemblages comply with narrative water quality standards. The Court cannot agree. Both WVSCI and GLIMPSS are valid and accepted methodologies to test for impairment. Moreover, expert testimony and peer-reviewed studies have shown that GLIMPSS is a more accurate measure of

[2] WVDEP and EPA jointly developed GLIMPSS in 2010. Letter from Shawn M. Garvin, Reg'l Adm'r, EPA Region III, to Randy C. Huffman, Sec'y, West Virginia Dep't of Envtl. Prot., Enclosure 2, 3 (May 11, 2016), Pls.' Ex. 79, ECF No. 72-30. It uses genus-level identifications of macroinvertebrates (insects, generally) collected at streams to calculate a score that corresponds to the health of the stream. *Id.*

[3] WVSCI is the predecessor to GLIMPSS and uses family-level identifications of macroinvertebrates to determine the health of a stream. Pls.' Ex. 79, Enclosure 1, at 11. It was developed by WVDEP and Tetra Tech, an environmental consulting firm, in 2000. *Id.*

impairment. Accordingly, the Court will not hamstring itself by limiting its review of the evidence to WVSCI.

WVDEP, in partnership with private consulting firm Tetra Tech, developed WVSCI in 2000 and began using it to determine impairment for purposes of compliance with the CWA in 2002.[4] Gregory J. Pond et al., *Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams*, 185 Envtl. Monitoring & Assessment 1515, 1516 (2013), Pls.' Ex. 122, ECF No. 72-34. WVSCI uses six metrics to assess assemblages of macroinvertebrates collected in the body of water targeted for study. *Id.* These metrics measure the presence of certain macroinvertebrates at the family level of taxonomic identification. *Id.* WVSCI, as well as other multi-metric indices, use macroinvertebrates to determine impairment because they "are typically considered . . . sensitive to anthropogenic disturbance" and "can quickly recolonize habitats under improving chemical or physical conditions." *Id.*

WVDEP and Tetra Tech developed WVSCI using 109 reference sites.[5] Reference sites are bodies of water that exhibit no, or at least minimal, anthropogenic influence. *Ohio Valley Envtl. Coal., Inc. v. McCarthy*, No. 15-cv-271, 2017 WL 600102, at *3 fn. 1 (S.D. W. Va. Feb. 14, 2017).

[4] In accord with 33 U.S.C. § 1313(d) states must periodically compile lists of bodies of water within their borders not meeting water quality standards and submit those lists to EPA for approval. These lists are known as "303(d)" lists. EPA has the authority to add bodies of water that it believes are not meeting water quality standards, including narrative water quality standards, based on a state's failure to evaluate existing and readily available data indicating that a body of water is not meeting water quality standards. *See* 40 C.F.R. § 130.7(b)(5).

[5] Tetra Tech questioned the quality of some of the reference sites included in the original sample used to develop WVSCI. Tetra Tech, A Stream Condition Index for West Virginia Wadeable Streams 23, Def.'s Ex. 45, ECF No. 73-7. It noted that some scored lower than a 68, which might indicate that they were misidentified as reference sites. *Id.* Upon further review, Tetra Tech opined, if a reference site revealed previously unidentified human influence, it should be discarded from the references site sample. *Id.*

Reference sites are used to define benchmarks for chemical, biological, and habitat conditions. *Id.* Macroinvertebrate assemblages collected and analyzed using WVSCI are assigned a score representing the target stream's similarity to the set of reference sites. WVSCI is scaled from 0 (worst) to 100 (best). *Id.* at *3. A score of 68.0 or below indicates impairment. Pls.' Ex. 79, at 1521. Developers of WVSCI determined that a score of 68.0 corresponded to the fifth centile of the 109 reference streams used to develop the index. Def.'s Ex. 45, at 21. In other words, ninety-five percent of all reference sites had a higher score. Pls.' Ex. 79, Enclosure 2, at 3. Setting the threshold as a percentile of reference site scores corresponds to limiting the chance that a healthy stream is found to be impaired, i.e. a false positive, to five percent. *Id.*

"Since publication of WVSCI in 2000, however, available biological data and science have progressed significantly." Pls.' Ex. 79, Enclosure 1, at 11. Assessment of the health of bodies of water has progressed from family-level to genus-level identification to "more accurately represent the composition of the aquatic community and increase[] [the] ability to detect a variety of impacts." Pls.' Ex. 122, at 1516. Regarding genus-level indices compared to their family-level predecessors in southern West Virginia waters affected by mining, genus-level metrics have been found to detect impacts more effectively than WVSCI. *Id.* This is in part because a site may lose several genera before an entire family was extirpated and therefore became visible to a family-level index. Pls.' Ex. 79, Enclosure 1, at 11–12. "For example, in a recent study, sample identification at the genus level taxonomy demonstrated loss of entire functional feeding groups. Loss of an entire functional feeding group (at the genus level) indicates ecosystem imbalance . . . ." *Id.* By not evaluating this genus-level data, EPA has explained, important information related to impairment may be missed. *Id.*

In 2010, at the request of WVDEP, EPA, alongside WVDEP, developed GLIMPSS—a genus-level index. Pls.' Ex. 79, Enclosure 2, at 2. In addition to calibrating GLIMPSS to the genus-level, GLIMPSS used nearly 400 reference sites to set natural undisturbed background benchmarks. *Id.* Using genus-level data collected for fourteen years by WVDEP but unused until GLIMPSS was developed, EPA and WVDEP identified distinct strata of data that corresponded both to seasonality and geography. *Id.* at 3; Pls.' Ex. 122, at 1522. The use of genus-level data revealed differences in the macroinvertebrate community that correlated with ecoregion (mountain or plateau) and season (spring or summer). Trial Tr. 154. Accordingly, GLIMPSS was calibrated using different threshold scores for each of the four stratum to capture the observed differences in four region-season classifications (mountain summer, mountain spring, plateau summer, plateau spring). Tr. 154; Pls.' Ex. 122, at 1532. Conversely, developers of WVSCI using family-level data "found no distinct natural classification patterns at the family-level, and thus the WVSCI is not tailored to geographic or seasonal variation . . . ." Pls.' Ex. 122, at 1516.

Like WVSCI, GLIMPSS was calibrated such that the threshold passing score was pegged to the fifth centile of the distribution of reference sites.[6] *Id.* at 1532. That is, ninety-five percent of all reference sites scored higher. The false positive rate therefore is still 0.05. Due to the seasonal and regional variation detected by GLIMPSS, each season-region strata has a threshold score pegged to the fifth centile of reference sites that correspond to that season-region. *Id.* For the

[6] Interestingly, in the peer-reviewed explanation of the development and calibration of GLIMPSS, the authors compared GLIMPSS to WVSCI in a number of metrics. Pls.' Ex. 122, at 1521. To do this the authors updated the WVSCI impairment threshold with all current reference sites (391 current versus 109 in the original WVSCI development). *Id.* The additional data moved the fifth centile of reference sites from a threshold of 68.0 to 71.6. *Id.* GLIMPSS was then used to score sites in each strata that had been scored by WVSCI. In the mountain summer strata (1530 sites) GLIMPSS assessed nine percent more impaired sites than WVSCI and ten percent more impaired sites in the mountain spring strata (697 sites). *Id.* at 1532.

mountain region, the region applicable to this litigation, a score of 53.0 or above is a passing score in spring, and 55.0 or above is a passing score in summer. *Id.;* Tr. 154.

The threshold score between impaired and unimpaired is not set arbitrarily. It represents the rate at which a stream that is in fact healthy is deemed impaired by the index. Pls.' Ex. 79, Enclosure 2, at 3. The threshold therefore embodies the quality of the reference sites used to develop the index, with higher quality reference sites permitting a lower threshold score and higher confidence that there are few false positives. It is not a "value judgment." *See id.*

For instance, other states that use indices to detect impairment have set thresholds at the tenth centile (e.g., Virginia), or even the twenty-fifth centile. *Id.* at fn. 7. The reference site sample in these other states lacked the same quality as that in West Virginia. *Id.* Reference sites that include more human influence mean background benchmarks that include more human influence. *See id.* Accordingly, the ability of the index to discriminate between impaired and unimpaired streams as measured by comparison with background conditions is diminished, resulting in a higher false positive rate. *See id.*

At the behest of WVDEP, the development and calibration of GLIMPSS was subjected to external peer review and published in a respected academic journal: Environmental Monitoring and Assessment. *Id.*; Tr. 153. In sum, as compared to WVSCI's family-level resolution, "[t]he fundamental improvement of genus-level data is representativeness. Compared with family-level taxonomic data, genus-level assessments more accurately represent the composition and diversity of the aquatic community in WV's flowing streams."[7] Pls.' Ex. 122, at 1533. Moreover, accuracy

[7] Another point made by the scientific literature is that indices like GLIMPSS or WVSCI act as representations of impairment. *See* Pls.' Ex. 122, at 1533. The rate at which an index tracks impairment defines its efficacy as a valid methodology to detect impairment. An index's ability to accurately reflect impairment is dependent on its ability to represent the health of all other inhabitants of a body of water, including fish, salamanders, mussels, and a host of other aquatic

in identifying impaired streams in West Virginia is bolstered by the inclusion of seasonal and regional specific calibrations. Tr. 154, 155; Pls.' Ex. 79, Enclosure 2, at 2.

Plaintiff's expert Dr. King, who is a professor of biology at Baylor University and teaches, conducts research, and publishes extensively on freshwater aquatic ecosystems, etymology, and ecological data analysis, testified that GLIMPSS is a more "rigorous" assessment tool and a better indicator of impairment than WVSCI. Pls.' Ex. 47; Tr. 159. Similarly, EPA has described GLIMPSS as more "modern" and "accurate" than WVSCI. Pls.' Ex. 79, Enclosure 2, at 2; Tr. 158. EPA also uses GLIMPSS to assess WVDEP's 303(d) List submissions. Pls.' Ex. 79. WVDEP urges the use of GLIMPSS where genus-level data is available. West Virginia Dep't of Envtl. Prot., Watershed Assessment Branch 2015 Field Sampling Standard Operating Procedures 5-43, Pls.' Ex. 133, ECF No. 72-35. Fola did not challenge the assessment of EPA, peer reviewed journals, or Plaintiffs' expert on the relative accuracy of GLIMPSS compared to WVSCI.

Fola's defense in this case has focused solely on WVSCI as *the* measure of impairment. Not without some irony, the Court feels compelled to note that in the *Stillhouse* litigation, Fola attempted to undermine the Court's reliance on WVSCI as a measurement of impairment. It complained to the Fourth Circuit that the Court's reliance on WVSCI scores to determine impairment "usurped [WVDEP's] role in its use of [WVSCI]" because "WVDEP . . . recently rejected [WVSCI] as the *sole* determinant of water quality."[8] *Fola*, 845 F.3d at 145 (emphasis in

---

life, by assessing a representative biological community (macroinvertebrates) as a stand-in for all those aquatic organisms. *Id.* In a sense, GLIMPSS looks for the "canary in the coal mine" and assesses the canary's health more accurately than WVSCI. *See Stillhouse*, 82 F. Supp. 3d at 698.

[8] In 2012 the West Virginia Legislature passed Senate Bill 562 ("SB 562"). *McCarthy*, 2017 WL 600102, at *6. The legislation requires WVDEP to develop a new methodology to test for compliance with the narrative water quality standards. *Id.* WVDEP thus stopped using WVSCI to measure impairment in 2012, although not without significant exceptions. *Id.* at *6, *7. WVDEP has not adopted GLIMPSS as the replacement for WVSCI but has proposed no other alternative to WVSCI or GLIMPSS. *Id.*; Pls.' Ex. 79.

original). Fola now wishes to bind the Court to WVSCI to the exclusion of all other scientifically relevant methodologies. On this point the Fourth Circuit held, "The [C]ourt did not enshrine [WVSCI] as the sole acceptable method of establishing violations of water quality standards." *Id.* at 146. The court went on to explain that in the absence of any "meaningful alternative" the Court's use of WVSCI was justified. *Id.*

Fola's current position that this Court cannot use anything other than WVSCI lest it usurp WVDEP's "value judgment" regarding what score constitutes a violation of the narrative water quality standards is self-serving and without merit. A determination of whether Fola's discharges impaired a receiving body of water is an issue of fact to be determined on a case-by-case basis upon specific evidence that some biological or physical attribute of the receiving stream has been degraded. *See Stillhouse*, 82 F. Supp. 3d at 696–98. As this Court has held on at least two occasions, and now with support from the Fourth Circuit, "violations of narrative water quality standards must be determined based on a reasoned and meaningful methodology." *Id.* at 679 (citing *Ohio Valley Envtl. Coal. v. Elk Run Coal Co., Inc.*, 24 F. Supp. 3d 532, 548–50 (S.D. W. Va. 2014)); *Fola*, 845 F.3d at 145. Whether Fola's discharges cause or materially contribute to: "[m]aterials in concentrations which are harmful, hazardous or toxic to man, animal or aquatic life; [a]ny other condition . . . which adversely alters the integrity of the waters of the State;" or "significant adverse impact[s] to the chemical, physical, hydrologic, or biological components of aquatic ecosystems" can be shown by any reasoned and meaningful methodology. *See Fola*, 845 F.3d at 145.

More to the point, a WVSCI or a GLIMPSS score represents a particular assemblage of macroinvertebrates that are representative of the health of all the other living creatures in a particular aquatic environment. The score, and the means by which it is produced, is shorthand for

the level of degradation in a stream. A failing score demonstrates that some biological or physical component of the stream has been degraded to the point that certain animals that would naturally occur in that stream are no longer present or have substantially diminished populations. This is all that needs to be shown to demonstrate that a body of water is not complying with narrative water quality standards. *See* W. Va. Code R. § 47-2-3.2.e, -3.2.i.

Multi-metric tools quantify the point at which impairment under the narrative standard occurs. Policy makers do not set the particular score that divides impaired from unimpaired, as Fola believes; scientists who developed the indices set the threshold. Thus, the value of the threshold score is a matter of scientific inquiry. The peer-reviewed journal that published the calibration of GLIMPSS explained that the GLIMPSS threshold was set at the fifth centile of reference sites due to the high quality of West Virginia's reference sites. Researchers had high confidence that human influences were minimal or nonexistent at these sites. With that confidence WVDEP and EPA scientists determined that a threshold of the fifth centile was supported by the data. In other words, the likelihood that GLIMPSS identified a body of water as impaired when it was not actually impaired is five percent. That is the same error rate as WVSCI, albeit using a far more robust reference sample.

The change that Fola fears—that use of something other than WVSCI alters the attributes that constitute impairment as defined by the narrative standard—is unfounded. GLIMPSS only assesses impairment more accurately. It is more accurate for a number of reasons, including a larger reference site sample that provided more information about background natural conditions, inclusion of seasonal and regional differences in assemblages, and, not least, the ability to analyze assemblages at the genus-level. None of these refinements change the narrative water quality standards. They only make it possible for GLIMPSS to better detect important changes in stream

health that were previously overlooked by WVSCI. Thus, a stream that is impaired but escaped identification when using WVSCI is still impaired.

Use of WVSCI or GLIMPSS in this case does not invade WVDEP's ability to choose how it would like to measure impairment for the purposes of complying with the CWA. WVDEP can choose how it will fulfill its duties however it wishes, subject to approval from EPA. *See* 33 U.S.C. § 1313. For the purposes of Fola's compliance, however, so long as the method presented to the Court is efficacious, accurate, reliable, and reasoned, as demonstrated by competent scientific expert testimony, the Court as well as regulatory agencies may use it to determine violations of the narrative water quality standards. Use by WVDEP or EPA confirms the attributes that a valid methodology must have. In general, the Court trusts that neither WVDEP nor EPA would use inaccurate or unreliable methods to determine compliance with water quality standards.[9] But, of course, this is not the only way to prove a methodology is valid. Accordingly, the Court will review both WVSCI and GLIMPSS scores submitted into evidence in this case.

B. GLIMPSS and WVSCI Scores

Both Plaintiffs and Fola calculated WVSCI and GLIMPSS scores for Elick Hollow and Shanty Branch. Sampling in each stream took place in October 2015 and April and August 2016. Pls.' Ex. 11, ECF No. 72-2; Pls.' Ex. 12, ECF No. 72-3. Plaintiffs conducted sampling in all three

[9] WVDEP and EPA used WVSCI for about a decade (2002 to 2010). *McCarthy*, 2017 WL 600102, at *3. Although WVDEP has not formally adopted GLIMPSS as the methodology it will use to determine which bodies of water to list on its 303(d) List, EPA uses GLIMPSS to review WVDEP's 303(d) List to determine whether WVDEP has used all available data in constructing its 303(d) List. Pls.' Ex. 79. As of May 2016 EPA used GLIMPSS to identify 61 impaired streams not included on WVDEP's 303(d) List. *Id.* Moreover, WVDEP also promotes the use of GLIMPSS in its Benthic Macroinvertebrate Collection Protocols, stressing that WVSCI should only be used for data that does not include genus level identifications. Pls.' Ex. 133, at 5-43.

months. *Id.* Fola only sampled in April and August 2016. Fola also only calculated GLIMPSS scores for its April 2016 sampling event. *Id.*

Site	Date	GLIMPSS	WVSCI
Elick Hollow	Oct. 2015	33.00	57.90
	Apr. 2016	46.60	75.90
	Apr. 2016 (Fola)	27.80	65.60
	Aug. 2016 (Fola)	N/A	61.65
	Aug. 2016	26.60	52.50
Shanty Branch	Oct. 2015	40.00	69.40
	Apr. 2016	17.20	58.00
	Apr. 2016 (Fola)	27.50	66.90
	Aug. 2016 (Fola)	N/A	59.86
	Aug. 2016	36.20	58.30

Neither stream registered a passing GLIMPSS score out of four sampling events for each stream. *Id.* Each stream also registered one passing WVSCI score out of five sampling events for each stream. *Id.* In April 2016 Elick Hollow received a WVSCI score of 75.90 and in October 2015 Shanty Branch received a WVSCI score of 69.40.[10] Based on this data, the Court concludes that both Elick Hollow and Shanty Branch are impaired and do not meet West Virginia's narrative

[10] Plaintiffs believe that the passing score recorded in Shanty Branch is a result of macroinvertebrates that were inadvertently added to Shanty Branch from an unimpaired stream by Fola. Def.'s Resp. to Pls.' First Set of Interrogs. and Second Req. for Produc. of Docs., Pls.' Ex. 52, ECF No. 72-25. Fola deposited 50 to 60 gallons of substrate from a healthy stream in Shanty Branch to remediate a spill of iron sludge resulting from the failure of the sludge curtain in the settling pond above Outlet 011. *Id.* The substrate may have contained macroinvertebrates, but there is no direct evidence that it did. At most the October Shanty Branch WVSCI score is questionable. Whether the introduction of macroinvertebrates occurred, whether they were then collected, and whether their presence in the full sample improved the resulting subsample is speculative. As the Court will explain more fully, a single passing score, even if valid, does not sway the Court from its determination that Shanty Branch is impaired. A conclusive finding of the effect of the "seeding" thus is not required.

water quality standards. The passing WVSCI score in each stream does not dissuade the Court from its conclusion. First, the more accurate measure, GLIMPSS, produced scores that uniformly failed. Second, a single passing WVSCI score out of five does not tip the balance in favor of Fola. As Justice Scalia artfully put it: "[a] good or lucky day is not a state of compliance." *Gwaltney of Smithfield, Ltd. v. Chesapeake Bay Found.*, 484 U.S. 49, 69 (1987) (Scalia, J., concurring). Failing four out of five times is enough to demonstrate by a preponderance of the evidence that both streams are impaired. The GLIMPSS scores corroborate this finding. Had the GLIMPSS scores been mixed, or had they been mostly passing, the Court may have doubted the failing trend in the WVSCI scores. That is not the case here. The GLIMPSS scores verify the impairment found by WVSCI.

C. Macroinvertebrate Assemblages

In addition to the WVSCI and GLIMPSS scores, the type and amount of macroinvertebrates found in Shanty Branch and Elick Hollow demonstrate serious adverse impacts to the natural macroinvertebrate community. Upon review of the types of macroinvertebrates found in the two streams represented in "taxa lists"[11] from each sampling event, Plaintiffs' expert Dr. Matt Baker[12] observed "[a]n unimpaired stream in this . . . part of the world would have many more taxa and many more genera and families." Tr. 129; *see also* Taxa List for October 9, 2015, Pls.' Ex. 13, ECF No. 72-4; Pls.' Taxa List for April 26, 2016, Pls.' Ex. 14, ECF No. 72-5; Def.'s Taxa List for April 26, 2016, Pls.' Ex. 15, ECF No. 72-6; Pls.' Taxa List for August 22, 2016,

[11] A taxa list is a list of the type and number of organisms used in the subsample to calculate an index score.

[12] Dr. Baker is a professor of geography and environmental systems at the University of Maryland, Baltimore County. Tr. 67. Dr. Baker specializes in environmental science, watershed science, geographic information systems, and riparian ecology. Tr. 67. Dr. Baker also teaches and publishes on issues related to the use of macroinvertebrates as indicators of the condition of freshwater systems and the effects of conductivity on freshwater systems. Tr. 68.

Pls.'s Ex. 16, ECF No. 72-7. He went on to note that in addition to the diversity of genera missing, "their abundances would be much greater." Tr. 129. In Dr. Baker's opinion, the small numbers of organisms collected on a few occasions by Dr. Swan, which will be discussed more fully in the next section, is further evidence that the streams are impaired. Tr. 132–33. He explained that "[n]ormally a healthy stream would be teeming with life. You would pick up rocks and leaves and there would be organisms everywhere . . . . And in these streams, . . . you have to look to find something." Tr. 133. In these streams, he noted, the organisms that were found were largely tolerant to high levels of conductivity. Tr. 129.

Reviewing the same taxa lists Plaintiffs' expert Dr. King held a similar view to that of Dr. Baker. Dr. King found that "looking at the taxa that are at the site now, they're highly consistent with what you'd expect to see as sites that are impaired by conductivity associated with surface mining." Tr. 163. He noted that the organisms that were found were "largely very tolerant organisms." Tr. 163. Dr. King also found the near total absence of mayflies in both streams to be indicative of impairment caused by conductivity. Tr. 164–65. Normally, he explained, mayflies constitute fifty percent of a subsample and sometimes up to ninety percent of a subsample. Tr. 189. Ultimately, Dr. King concluded that Shanty Branch and Elick Hollow were "[d]efinitely not" healthy Appalachian streams. Tr. 165. In addition, a near complete absence of mayflies in Stillhouse Branch coupled with failing WVSCI scores led this Court to determine that it too was impaired. *Stillhouse*, 82 F. Supp. 3d at 697.

The array of evidence before the Court here necessitates the same finding. Shanty Branch and Elick Hollow are impaired. The WVSCI and GLIMPSS scores show consistent impairment, which is confirmed by the reduced abundance and diversity of macroinvertebrates in both streams.

D. Sampling Methods

In its defense, Fola maintains that the failing WVSCI scores from Elick Hollow in October 2015 and Shanty Branch in April 2016 were procured by Plaintiffs' expert with unorthodox methods inconsistent with the protocols set by the WVDEP for collecting WVSCI scores. Fola did not argue that sampling methods affected any of the GLIMPSS scores.

Dr. Christopher Swan collected the macroinvertebrate samples used by Plaintiffs to calculate WVSCI scores. Dr. Swan is a professor of geography and environmental systems at the University of Maryland, Baltimore County. Tr. 12–13; Pls.' Ex. 8. In that capacity Dr. Swan teaches and performs research on stream ecosystems and macroinvertebrates and teaches macroinvertebrate sampling techniques. Tr. 13–14. Dr. Swan has also published extensive research based on macroinvertebrate samples. Tr. 13. Dr. Swan has used the West Virginia macroinvertebrate sampling protocols at least fifteen times and has testified about the results in court on a number of occasions. Tr. 15.

To sample a stream Dr. Swan first measures out a 100-meter segment of the stream where possible. Tr. 16. He then looks for available habitat throughout the 100-meter segment. Tr. 16–17. He attempts to sample from habitat distributed roughly equally throughout the segment. Tr. 17. Dr. Swan explained that riffles are the target habitat to be sampled.[13] Tr. 17. The preliminary steps taken by Dr. Swan before sampling comply with the protocols established by WVDEP. *See* Pls.' Ex. 133, at 5-5.

Once habitat has been located throughout the segment, Dr. Swan places the kick-net on the bottom of the channel in the middle of stream and downstream from the habitat to be sampled. Tr.

[13] Riffle habitat is described by WVDEP as "[w]ater moving with small ripples, waves and eddies. Produced a babbling or gurgling sound." Pls.' Ex. 133, at 5-5.

17. He then kicks his feet while counting to thirty in his head to disturb the bottom of the stream so that any animals are lifted up by the current and transported to the net for collection. Tr. 17–18, 23. Dr. Swan testified that he always counts to thirty in his head while kicking. Tr. 23.

The WVDEP protocol requires the sampler to kick for twenty seconds down to a depth of ten centimeters. Pls.' Ex. 133, at 5-8. The WVDEP protocol also explains that before kicking, surfaces of cobble and boulders should be brushed to wash any animals into the net. *Id.* The rocks should be removed if possible and the undersides brushed so that any animals clinging to them are collected in the net. *Id.*

Dr. Swan testified that he kicks first and then uses his hand to brush off any large substrate. Tr. 22. He then visually inspects the substrate to ensure that he did not miss any macroinvertebrates. Tr. 22. Dr. Swan believes that using his hands instead of a bristle brush results in a better sample because a bristle brush can damage the morphology of some macroinvertebrates. Tr. 22.

WVDEP protocol further notes that from the sample collected at the site, 200 organisms plus or minus twenty percent are used to determine a WVSCI score. Pls.' Ex. 133, at 5-43; Tr. 132–33. Where fewer than 100 organisms are collected, "[a]s a rule of thumb," those samples "should be carefully scrutinized for comparability before calculating a WVSCI score." Pls.' Ex. 133, at 5-43. The protocol explains that sites yielding fewer than 100 organisms "may be heavily impacted by stressors or were recently subjected to drought or scour events." *Id.* On this last possibility, the Court will have much to say later.

Dr. Swan produced two samples under 100 organisms. In April 2016, Dr. Swan's sampling of Shanty Branch yielded only 85 organisms. Taxa List for April 26, 2016, Pls.' Ex. 14, ECF No. 72-5. The WVSCI score calculated for Shanty Branch by Plaintiffs on that day was 58.00. In

August 2016, again in Shanty Branch, Dr. Swan's sampling yielded only 87 organisms. Taxa List for August 22, 2016, Pls.'s Ex. 16, ECF No. 72-7. The WVSCI score calculated by Plaintiffs for Shanty Branch on that day was 58.30. Fola does not argue that sampling errors led to the low numbers in August 2016. It advances a different argument that the Court will address in the next section.

Fola hired Ray Ewing, an employee of environmental consulting firm Civil Environmental Consultants ("CEC"), to observe Dr. Swan's sampling techniques. Tr. 202. Fola presented Ewing as a fact witness albeit with experience conducting benthic macroinvertebrate sampling through his employment with CEC. Tr. 198–200. Ewing testified that he is familiar with WVSCI protocols as set out by WVDEP. Tr. 201. He also explained that he has conducted hundreds of samples himself. Tr. 202. Ewing was on hand to observe Dr. Swan's technique during the October 2015 sampling event. In Elick Hollow Ewing testified that he did not see Dr. Swan brush off large rocks before kicking. Tr. 217. Rather, Dr. Swan "did more of a rinsing motion" on his second kick, no rinsing or brushing on his third kick, and no rinsing or brushing on his fourth kick. Tr. 217. Ewing did not observe Dr. Swan's first kick in Elick Hollow. Tr. 216. Ewing inspected the streams and saw cobble and larger material "in most of them." Tr. 217. Ewing did not render any testimony concerning whether Dr. Swan brushed off any substrate in the two other streams.

Ewing also timed Dr. Swan's kicks in Elick Hollow by counting in his head. Tr. 216. He missed the first kick, but estimated that Dr. Swan kicked for eleven seconds, twenty seconds, and twelve seconds for the second, third, and fourth kicks, respectively. Tr. 215–16. Plaintiffs calculated a WVSCI score of 57.90 for Elick Hollow in October.

In April 2016, Ewing observed Dr. Swan when he sampled Shanty Branch. Tr. 218. Ewing testified that he did not see Dr. Swan brush off large rocks before kicking and only brushed off

larger substrate after he kicked if it was in the net. Tr. 219. Ewing also estimated three out of four of Dr. Swan's kicks in Shanty Branch at nineteen seconds, seventeen seconds, and sixteen seconds, respectively. Tr. 219. Dr. Swan collected 85 organisms in Shanty Branch in April 2016. Plaintiffs calculated a WVSCI score of 58.00 in Shanty Branch in April.

Fola maintains that the Court should discount Dr. Swan's WVSCI scores from Elick Hollow in October 2015 (57.90) and Shanty Branch in April 2016 (58.00) due to sampling protocol defects. Fola insists that Dr. Swan's failure to comply with WVDEP protocols depressed the two WVSCI scores, and without them, there is no conclusive evidence of impairment in either stream. Fola adduced no expert testimony explaining the relationship between sampling technique and WVSCI scores. The inference intended by Fola is that where protocol is not followed so few organisms are collected that the WVSCI score is depressed. The evidence and testimony presented at trial does not support this inference, at least not to the degree that would make a difference to the outcome of this case.[14]

Due to the way WVSCI and GLIMPSS are calculated, the total number of organisms collected at the site is unknown, unless the total number falls below 160. Pls.' Ex. 133, at 5-17 to 5-20. If the total number is above 160, the number of organisms used to calculate the score is a subsample of the total number arrived at through random sampling of organisms distributed over a numbered grid. *Id.* The goal of subsampling is to get the subsample as close to 200 organisms as possible without exceeding 240 or falling below 160. *Id.* Each grid space may contain no

[14] WVSCI scores are not correlated to the number of organisms collected. Tr. 112 (testimony of Dr. Baker explaining the six metrics used to calculate WVSCI. None are total number of organisms collected). WVSCI scores are much more sensitive to the kind of macroinvertebrates collected and the numbers of certain families of macroinvertebrates. Tr. 113. The only concern raised by WVDEP is for samples under 100 organisms, and even then the sample is not disqualified. Pls.' Ex. 133, at 5-43.

organisms or may contain multiple organisms, and the subsampler continues to randomly select grid spaces until a subsample as close to 200 as possible is reached after counting all organisms in a the selected grid space. *Id.* Thus, a subsample that contained 193 organisms came from a full sample of 193 organisms *or more*. *See id.* But, a subsample of 87 organisms came from a full sample of 87 organisms. *See id.* In the latter scenario, all the organisms present in the full sample were selected in an attempt to reach the 200 organism goal. *See id.* This is true for any full sample containing fewer than 160 organisms (200 minus twenty percent). *Id.* Subsamples below 160 occurred in both Fola and Plaintiffs' samples in all months. Pls.' Exs. 13, 14, 15, 16, 18, ECF Nos. 72-4, 72-5, 72-6, 72-7, 72-8. WVDEP protocol however only raises concern over samples that fall below 100 organisms. *See* Pls.' Ex. 133, at 5-43. Samples yielding 100 and 160 do not raise concerns about the reliability of the sample, according to WVDEP. *See id.*

Taking the October 2015 Elick Hollow sample first. The Court would note that Dr. Swan collected enough organisms to yield a subsample of 193. Pls.' Ex. 13. By contrast, Dr. Swan collected only 150 organisms in Shanty Branch in October, presumably using the same techniques, and calculated a passing WVSCI score of 69.40. *Id.* In April 2016, Dr. Swan collected enough organisms to yield a subsample of 178. Pls.' Ex. 14. On that same day Fola's team collected a full sample of only 125 organisms. Pls.' Ex. 15. It appears that Dr. Swan's sampling technique, compliant or not, cannot explain the data produced at trial and did not reduce the number of organisms in any way that would have made calculating a WVSCI score questionable under current WVDEP protocols.

The Court further credits Dr. Swan's testimony that he counts to thirty every time he samples. Dr. Swan is an expert in the field, conducts his own macroinvertebrate testing to use in peer-reviewed research, and teaches sampling techniques. Moreover, Plaintiffs' expert Dr. Matt

Baker observed Dr. Swan during the April outing and testified that Dr. Swan was using standard sampling protocol and "if the stream had a lot of organisms, they would be obtained by the way that he sampled." Tr. 135. Dr. Swan testified that he did not change his technique between October and April. This is not to say that Ewing fabricated his testimony. Rather, Dr. Swan's testimony should not be discarded because of a mental estimation of time.

Lastly, WVSCI and GLIMPSS scores calculated in April and August are not markedly different from the scores calculated in October. They largely confirm the October scores. Had Dr. Swan's methods affected WVSCI scores (Fola makes no argument directed at the GLIMPSS scores), the Court would expect to see marked difference when Fola conducted its own sampling. That is not the case.

Turning to the April 2016 WVSCI score from Shanty Branch, Fola again takes issue with Dr. Swan's sampling techniques and noted that he collected fewer than 100 organisms. As Ewing observed, and Dr. Swan testified, he did not remove large substrate before kicking in Shanty Branch. Ewing also estimated that Dr. Swan kicked for less than twenty seconds on three out of four of his kicks in Shanty Branch. Fola asserts these departures from protocol depressed the WVSCI score for Shanty Branch, making it an unreliable measure of the health of the stream. Without some expert testimony connecting the relationship between certain protocols and WVSCI scores, the Court is unwilling to accept the inference that failure to strictly observe certain protocols produced an inaccurate unidirectional decrease in WVSCI scores.[15]

Moreover, unlike the October 2015 Elick Hollow WVSCI score where only Plaintiffs collected data, Fola collected its own samples and calculated both WVSCI and GLIMPSS scores

[15] Fola's expert opined that differences in sampling protocol "can cause [scores] to vary." Tr. 264. This is the extent of Fola's expert testimony concerning the relationship between sampling protocol and WVSCI scores.

for Shanty Branch in April. Fola in fact sampled on the same day as Plaintiffs only one hour after Plaintiffs collected their samples. Tr. 223–24. There is no question that Fola observed WVDEP sampling protocol. Fola calculated a WVSCI score of 66.90 (68.0 threshold) and a GLIMPSS score of 27.50 (53.0 mountain spring threshold).[16] Both are failing scores. This is enough to convince the Court that while the particular WVSCI score calculated by Plaintiffs may be somewhat less reliable than Fola's scores, Shanty Branch is impaired. All scores calculated by both parties fell below the thresholds used to determine resemblance to reference conditions. These scores show that it is highly likely that Shanty Branch has deviated significantly from background conditions and is therefore impaired. The degree of that deviation is immaterial to this case at the liability phase.[17]

E. August WVSCI Scores

In June 2016, the Monoc #2 mine, along with other parts of southern West Virginia experienced a large rain event. Tr. 269. Rain gauges at the Fola site registered six inches of rainfall over a 48-hour period. Tr. 270–71. Fola contends that the rain event in June artificially depressed WVSCI scores collected in August because the macroinvertebrate community was likely affected by the higher flows and probably did not recover by the August sampling. Fola called on Dr. Tim

[16] Developers of GLIMPSS calculated a confidence interval of ninety percent for single observations. Pls.' Ex. 122, at 1521.

[17] Fola's argument also lacks credibility. Although Fola expects the Court to infer that Dr. Swan's sampling method yielded fewer organisms than Fola's sampling methods and therefore lower WVSCI scores, as already explained, the evidence before the Court does not permit that inference. Based on the evidence before the Court, and were it to credit Fola's argument, it can at most infer that samples collected by Dr. Swan are unreliable. Yet, Fola does not attempt to invalidate passing WVSCI scores that resulted from Dr. Swann's allegedly improper methods, only the failing scores. Yet, Dr. Swann testified that he used the same techniques in October and April for all samples. Tr. 23. Were the Court to credit Fola's argument, both passing WVSCI scores (one in October and one in April) would be invalidated, leaving only failing WVSCI scores for both streams.

Verslycke to offer testimony about the rain event. Dr. Verslycke is currently employed as a principal at Gradient, an environmental consulting firm where he focuses on ecological risk and risk services. Tr. 237–38. Through his work at Gradient he directs research that incorporates the use of macroinvertebrate sampling and indices. Tr. 244. Dr. Verslycke obtained his doctorate in applied biological sciences from the Ghent University. Tr. 240. As part of his course work Dr. Verslycke studied freshwater and marine ecology and conducted research using macroinvertebrate indices, although he has never been to a West Virginia stream, never used WVSCI, and he did not address GLIMPSS in his testimony. Tr. 240–41, 261. Dr. Verslycke's published research generally addresses impacts of certain chemicals on marine macroinvertebrates with a specific focus on chemicals that affect endocrine systems. Def.'s Ex. 41, ECF No. 73-6.

Dr. Verslycke expressed that the WVSCI scores produced from the August sampling event might be unreliable due to the rain event that occurred two months prior. Tr. 268–69. He noted that in Shanty Branch in August, Dr. Swan collected only 87 organisms and this low number could be related to the lingering impacts of the June storm. Tr. 268. In addition, Dr. Verslycke explained that it is possible that a rain event can affect WVSCI scores, especially if the higher flows scoured the streambed. Tr. 268, 271. Nonetheless, he qualified this statement by noting that the impact and subsequent recovery of the stream is highly dependent on, among other things, the size of the storm, the specific site, the type of stream, and the time of year. Tr. 271. He did not express an opinion about how this specific storm affected these specific streams. Tr. 277. In fact, he did not do a site visit. Tr. 276. He restricted his opinions to explaining the possibility of an effect from the storm. Tr. 277–78. He went on to testify that recovery of Shanty Branch and Elick Hollow may take longer than a natural stream because after a storm organisms recolonize from nearby habitat, including drifting downstream, but here all but the lower reaches of the streams are either valley

fill or settling ponds—not habitat from which organisms could launch a recolonization. Tr. 273–74.

The Court has no doubt that a rain even can affect the abundance of macroinvertebrates in streams, generally. Signs that the June rain event affected the August WVSCI scores are conspicuously absent here. Fola's expert opined that the collection of 87 organisms in Shanty Branch in August was a "red flag" and that the paltry number of organisms might be related to the June storm. Tr. 268. None of the other samples taken that day by either party produced a sample with fewer than 100 organisms. Pls.' Ex. 16; Pls.' Ex. 60. In fact, Fola's sample from Shanty Branch just one hour later collected 131 organisms.[18] Pls.' Ex. 60. The other samples from that day included enough organisms to meet the 200 plus or minus twenty percent goal for sample size. Pls.' Ex. 16; Pls.' Ex. 60. Had the storm washed away large numbers of macroinvertebrates, and had they failed to recolonize the streams, the Court would expect collecting a sufficient number of organisms to be a challenge in all three streams.

Dr. Verslycke also explained that the variability in the WVSCI scores between those calculated in October and April and those in August might be explained by the June rain event. This explanation does not fit the facts presented here. WVSCI scores for Elick Hollow, save one, are remarkably similar. Four out of the five scores fall between 52 and 66. The April 2016 score of 75.90 appears to be an outlier. Indeed, on the day in April when that score was recorded, Fola recorded a score of 65.60. When compared with the GLIMPSS scores in Elick Hollow, there is no significant change between the scores recorded in October (33.00), April (46.60 & 27.80), and August (26.60). The same is true for Shanty Branch. All WVSCI scores for Shanty Branch are

[18] Collecting fewer than 160 organisms does not appear to be exclusively related to the June storm. Fola collected only 125 organisms from Elick Hollow in April. Pls.' Ex. 15.

between 70 and 58 with no great reduction in scores between April (58.00 & 66.90) and August (58.30 & 59.86). The GLIMPSS scores for Shanty Branch in August (36.20) again show no impact from an anomalous event when compared with earlier scores (April: 17.20 & 27.50; October: 40.00).

Plaintiffs' experts did not believe that signs of the June storm appeared in the August data. Both Dr. Swan and Dr. Baker explained that in their expert opinion waiting one month after a rainstorm insures that macroinvertebrate sampling is not affected by the storm. Tr. 19, 349–50. The parties waited two months after the June storm to sample again. Although Fola challenged Plaintiffs' experts' one-month assumption with research indicating that full recovery of a stream may take longer, Fola has not been able to convincingly explain the data collected from these specific streams. As Dr. Baker explained when he is concerned with the effects of a rain event on a sample, he completes the sample and then looks to see if it is in some way anomalous from what would be expected. Tr. 351. There are no anomalies in the data presented to the Court that correspond with a rain event. WVSCI scores did not collapse in August, nor were the streams passing in April. The GLIMPSS scores tell the same story—consistently, albeit not dramatically, impaired streams.

When Dr. Swan visited the site to take August samples, he was informed that there was a large storm in June but he "was surprised to learn that it was that magnitude . . . when I arrived at the head of the streams because I didn't see anything much different than I had seen in the past in terms of channel movement." Tr. 19. Dr. Verslycke did not visit the sites to confirm his hypothesis and thus his testimony, although informative, carries less weight than that of Dr. Swan.

It is certainly the case that rain events of a certain magnitude can cause macroinvertebrate populations to decline and thus might affect WVSCI scores. It is also true that the streams at issue

here were subject to a rainstorm of considerable magnitude. The effect of that storm, however, does not make itself known in the data collected by both parties. WVSCI and GLIMPSS scores were not depressed compared with their pre-storm counterparts. Both Dr. Swan and Dr. Baker testified that, in their expert opinions and in reliance on their experience sampling in West Virginia streams and similar habitats, one month is generally enough time to allow after a storm to get a sample unaffected by a storm. Finally, Dr. Swan explained when he visited in August he did not observe any significant changes to the stream that one might expect from a large rain event.

F. Cause of Impairment

Fola's Monoc #2 mine is the only significant contributor to Outfalls 005 and 011. Pls.' Ex. 51 ¶¶ 2, 3; Tr. 85. Those outfalls discharge directly into Shanty Branch and Elick Hollow. Pls.' Ex. 51. Discharge flows consistently throughout the year and contains high levels of conductivity as well as high levels of certain ions known to increase conductivity in Appalachian streams subject to mine drainage. Pls.' Ex. 51 ¶¶ 8, 9. There is no other significant source of flow to Shanty Branch or Elick Hollow.[19] Tr. 85, 88. Dr. Baker found that the watersheds for each stream were completely dominated by the Fola mine. Tr. 72.

The parties stipulated to the results of water sample testing done before Fola began mining operations up to the present both instream and at the discharge points. Pls.' Ex. 51 ¶ 7. As already noted conductivity levels in the two streams before mining was quite low and representative of natural background conductivity in Appalachian streams. Tr. 74. Before mining began, Fola reported that conductivity in Elick Hollow measured 35 µS/cm, while conductivity in Shanty Branch measured 44 µS/cm. Pls.' Ex. 51 ¶ 7. Sampling in 2015 and 2016 found conductivity in

[19] Dr. Baker noted that a small seep contributed to Elick Hollow but it was not a significant source of discharge into the stream. Tr. 85.

Elick Hollow at 2,060 µS/cm and 1,854 µS/cm, respectively. Summ. of chem. Sampling results by Evan Hansen, Pls.' Ex. 24. Conductivity in Shanty Branch for the same period measured 1,860 µS/cm. *Id.* Conductivity from Outfalls 005 and 011 measured 1,897 µS/cm 1,861 µS/cm, respectively for the same period. *Id.* The increase in conductivity has been tracked by increases in other compounds known to be the causative ions of conductivity in Appalachian mine drainage. *See id.*

Sulfate levels before mining were also much lower than those recorded after mining began. Levels in Elick Hollow measured 9.88 mg/L. *Id.* At the same time, Fola reported sulfate levels in Shanty Branch at 23.04 mg/L. *Id.* Sampling from 2015 and 2016 revealed sulfate levels of 1,100 mg/L in both Shanty Branch and Elick Hollow. Pls.' Ex. 24. Sulfate levels in the discharge from Outfalls 005 and 011 measured between 1,000 mg/L and 1,100 mg/L over the same period. *Id.* Similarly, alkalinity before mining began was measured at 9.54 mg/L in Elick Hollow and 5.30 mg/L in Shanty Branch. Pls.' Ex. 51 ¶ 7. Sampling from 2015 and 2016 found instream alkalinity levels ranging between 25 mg/L and 44 mg/L in Elick Hollow and between 64 mg/L and 67 mg/L in Shanty Branch. Pls.' Ex. 24. The alkalinity of the discharge from Outfalls 005 and 011 was nearly identical to that found instream at both streams. *Id.*

Dr. Baker also explained that other ions, along with sulfates and alkalinity, consistent with high levels of conductivity in Appalachian mine drainage were elevated compared to natural background conditions. Tr. 81. These additional ions are magnesium and calcium. Tr. 81. Fola did not report magnesium and calcium levels before mining began but it did not question Dr. Baker's assessment that they are elevated compared to background natural conditions. *See* Pls.' Ex. 24. Taking magnesium and calcium along with sulfates and alkalinity "[t]hey are the same ions that

are associated with alkaline mine drainage that the [scientific] literature has linked to degradation of macroinvertebrate assemblages." Tr. 81.

The scientific literature supporting Dr. Baker's finding was extensively reviewed by this Court in two prior cases related to highly conductive discharges of mine drainage. *See Stillhouse*, 82 F. Supp. 3d at 687; *Leatherwood*, 120 F. Supp. 3d at 544–46. These two cases were the bases of this Court's Memorandum Opinion and Order granting partial summary judgment on Plaintiffs' collateral estoppel on general causation. ECF No. 74. The Court's earlier opinions found that the large amounts of scientific data, including highly respected research from EPA, supported a finding that conductivity above 300 µS/cm and composed of elevated levels of alkalinity, sulfates, magnesium, and calcium from mine drainage causes degradation of macroinvertebrate assemblages in Appalachian streams. *Stillhouse*, 82 F. Supp. 3d at 686–95 (finding "[a]s conductivity increases, the occurrence and capture probability decreases for many genera in West Virginia at the conductivity levels predicted to cause effects. The loss of these genera is a severe and clear effect.").

This general proposition has been further bolstered by recent experimental confirmation. Plaintiffs' expert Dr. King explained that researchers were able to construct artificial streams large enough to replicate an Appalachian stream ecosystem where they were able to control for confounding variables such as habitat quality and other aspects of water quality. Tr. 165–66; William H. Clements & Chris Kotalik, *Effects of major ions on natural benthic communities: an experimental assessment of the US Environmental Protection Agency aquatic life benchmark for conductivity*, 35 Freshwater Science 126, 126 (2016), Pls.' Ex. 85, ECF No. 72-32. The researches then dosed the streams with differing levels of ions and found that conductance levels "near or less than 300 µS/cm . . . [affected] mayfly drift, abundance of baetid and heptageniid mayfly

abundance, and community metabolism." Pls.' Ex. 85; Tr. 166. Dr. King noted that it was "phenomenal" to have experimental confirmation of field findings and EPA's benchmark threshold of 300 µS/cm. Tr. 166.

The signature of impairment from conductivity was also present in the types of organisms collected during the parties' sampling events. Dr. Baker reviewed the "taxa lists" for each sampling event. After reviewing the lists Dr. Baker found that there were almost no mayflies at all and the caddisflies that were found were from taxa that are "exceptionally tolerant" to high levels of conductivity. Tr. 102, 107. Dr. King also reviewed taxa lists from the sampling events and came to the same conclusion—the presence of largely very conductivity tolerant organisms. Tr. 163. He found almost no mayflies in the taxa lists but explained that it is "improbable in a natural Appalachian stream . . . mayflies are very common, very diverse in healthy Appalachian streams."[20] Tr. 164. "It's not unusual," Dr. King explained, "for the percent of the total [mayflies] in a 200-count sample to be over 50 percent . . . and sometimes 80, 90 percent of the individuals will be mayflies." Tr. 189. Instead, the taxa lists overwhelmingly contained conductivity tolerant taxa. Tr. 189. A similar array of taxa led this Court to conclude that conductivity from mine drainage impaired Stillhouse Branch. *See Stillhouse*, 82 F. Supp. 3d at 697 ("As ionic concentrations increase, mayflies . . . dramatically decline . . . . [T]he sample was dominated by conductivity tolerant macroinvertebrates.").

G. Confounding Factors

Outside of its criticism of Dr. Swan's sampling methods and the June rainstorm, Fola did not raise any other factors that could be responsible for the impaired nature of the streams.

[20] Fola pointed out on cross-examination of Dr. King that the Fola sample from Shanty Branch in April contained a single mayfly. Tr. 178.

Plaintiff's experts reviewed both the habitat at each stream and other water quality data and determined that no other feature of the streams as they are currently composed could be responsible for the impairment detected by WVSCI and GLIMPSS. Dr. Baker reviewed both pH and temperature in each stream. Tr. 74. The pH of Shanty Branch before mining was 5.39 to 6.34. Pls.' Ex. 51 ¶ 7. Presently pH is between 7.9 and 8.3. Pls.' Ex. 24. The pH of Elick Hollow before mining began was 5.85 to 6.12. Pls.' Ex. 51 ¶ 7. At present the pH in Elick Hollow is 6.4 to 7.0. Pls.' Ex. 24. Dr. Baker explained these pH levels are within the range expected in Appalachian streams. Tr. 74. This was true of the temperature as well. Tr. 74. The two streams have a temperature range between 16.2 and 17.7 degrees Celsius, which Dr. Baker noted was within the range of variability for all Appalachian streams, including reference sites and other unimpaired streams. Tr. 81, 93; Pls.' Ex. 24; Pls.' Ex. 30.

Dr. Baker's evaluation of the quality of the habitat at each stream was consistent with Fola's evaluation. Both parties found it to be of generally good quality. The parties used the Rapid Bioassessment Protocol ("RBP") to evaluate habitat quality. Tr. 86. The RBP has a number of scored categories and the scores are based on visual observations of the site. Tr. 86. The scores are then summed to arrive at a final score. Tr. 87. The final scores are then broken into four categories; optimal, suboptimal, marginal, and poor. Tr. 87. Habitat scores in the marginal or poor ranges can have a significant impact on macroinvertebrate assemblages. Tr. 87. There is little evidence that scores in the optimal and suboptimal ranges have any impact on macroinvertebrates. Tr. 87.

Dr. Baker observed at Elick Hollow that the canopy around the stream had been disturbed but regrowth had occurred. Tr. 86. He also noted that there was a road adjacent to the stream and it was contributing some erosion, which he found accumulated near the lower part of the reach. Tr. 86–87. He did not believe this affected sampling. Tr. 87. Due to the high quality riffle habitat

in Elick Hollow, however, the stream scored in the "suboptimal" range, "which generally indicates good condition." Tr. 87. Fola's team concluded the same. Tr. 88; Pls.' Ex. 60.

While evaluating Shanty Branch, Dr. Baker observed a similar configuration to Elick Hollow. Tr. 89. It had experienced some past disturbance due to low-grade agriculture and was slightly more incised than Elick Hollow. Tr. 89. At Shanty Branch, however, the surrounding habitat was in better condition, the road was farther from the stream, and it did not receive erosion from the road. Tr. 89. Dr. Baker scored Shanty Branch as suboptimal. Tr. 89. Fola's scores for Shanty branch largely tracked those of Dr. Baker's. Tr. 89. Considering the RBP scores for both streams, Dr. Baker concluded that "[w]hen you see impaired streams at that level, that would seem to indicate that factors other than habitat are causing impairment." Tr. 92. Dr. Baker further noted that the RBP scores "suggest[] that the habitat are of high quality, among the highest qualities you see in West Virginia, and that they're consistent with reference streams in the state." Tr. 93. Fola did not challenge this assessment.

Accordingly, although Plaintiffs need not rule out every other possible factor contributing to impairment, Plaintiffs have presented a significant amount of evidence disqualifying habitat, other features of water quality, and land uses other than Fola's mine that could have contributed to the impairment found by WVSCI, GLIMPSS, and the aberrant assemblage of macroinvertebrates found at both Shanty Branch and Elick Hollow. The Court has little doubt that conductivity is the cause of impairment in these two streams and that the source of conductivity found in Shanty Branch and Elick Hollow is Fola's Monoc #2 mining operation.

H. Leatherwood Creek

Plaintiffs have not met their burden to demonstrate that discharges from Outlets 005 and 011 have caused the impairment detected in Leatherwood Creek. As Fola points out, there are

numerous discharges from other surface mines into Leatherwood Creek upstream from Shanty Branch and Elick Hollow. Def.'s Ex. 65. Plaintiffs submitted no evidence quantifying the contributions, if any, of conductive discharges from upstream outlets compared with the contributions from Shanty Branch and Elick Hollow. Plaintiffs' sampling station was located downstream from both Shanty Branch and Elick Hollow. Def.'s Ex. 65. Plaintiffs did not have a sampling site upstream from either stream at issue here. *Id.* Accordingly, although it appears quite likely that Leatherwood Creek is impaired, Plaintiffs have not demonstrated by a preponderance of the evidence that discharges from Outlets 005 and 011 have caused or materially contributed to that impairment.

In Plaintiffs' reply brief, they request the Court permit it to amend its Complaint to conform to the evidence pursuant to Federal Rule of Civil Procedure 15(b)(2). Pls.' Post-Trial Reply Br. 9, ECF No. 81. Plaintiffs would like to remove their claims related to Leatherwood Creek. Rule 15(b)(2) permits amendments to conform to proof adduced at trial, even post judgment, "[w]hen an issue *not* raised by the pleadings is tried by the parties' express or implied consent . . . ." Fed. R. Civ. P. 15(b)(2) (emphasis added). The issue prompting Plaintiffs' request to amend is one they would like to delete from their pleadings. Clearly, the Federal Rules do not intend Rule 15 to be used as an eraser for unproven claims. The function of the Rule can be broadly stated as ensuring that the claims raised in the pleadings mirror the claims litigated at trial. *See Moore's Federal Practice* § 15.18 (3d ed. 2009). Here there is nothing to conform. Plaintiffs raised claims related to Leatherwood Creek in their Complaint and introduced evidence at trial related to those claims. They simply failed to clear their evidentiary bar with regard that stream.

I. April 2017 WVSCI Scores

Fola submitted along with its Post-Trial briefing a motion to supplement the record or for a new trial so that Fola could submit passing WVSCI scores that it calculated after sampling in April 2017. Def.'s Mot. to Suppl. the R. or in the Alternative for a New Trial, ECF No. 78. Trial was held in mid-March. The Court sees no reason to supplement the record or order a new trial. It is enough that at the time of trial Plaintiffs demonstrated that Fola was in violation of water quality standards incorporated into its NPDES permits. A contrary finding would jeopardize the finality of any NPDES violation case by inviting a constant contest to procure desired results. Cases would lurch along, propelled by each new finding to a different result. At some point the Court must put an end to evidence collection; that point is trial.

Fola had plenty of time to conduct sampling to support its defense before trial. It was not constrained by any sampling protocol after the parties concluded their sampling in August 2016. WVDEP permits WVSCI sampling until October 15. Pls.' Ex. 133, at 5-43. Fola had two additional months to sample if it wished. In addition, the sampling period for GLIMPSS is nearly year round—December 1 to October 15. *Id.* at 5-40. The Fourth Circuit has held that a party must, among other things, exercise due diligence in discovering any new evidence to support a new trial. *Boryan v. United States*, 884 F.2d 767, 771 (4th Cir. 1989). There is no reason known to the Court why Fola could not have done additional sampling before trial. Accordingly, Fola has not exercised due diligence in collecting new evidence and therefore has not demonstrated that a new trial is warranted.

III. Conclusion

Plaintiffs' have shown by a preponderance of the evidence that both Shanty Branch and Elick Hollow are impaired due to discharges of certain ions as measured by conductivity from

Outlets 005 and 011 in violation of its NPDES permits. The chemical and biological components of Shanty Branch and Elick Hollow have been dramatically affected by Fola's discharges into each stream. WVSCI and GLIMPSS scores calculated over a number of months consistently show Elick Hollow and Shanty Branch to be impaired. The water chemistry of both streams reveals extremely high levels of ionic salts—measured as conductivity—known to cause the extirpation of large segments of the naturally occurring macroinvertebrate community. The biological characteristics of the streams have also been significantly injured. In both streams nary a mayfly was found over multiple samples taken at different times of the year. The macroinvertebrates that were present were largely, if not exclusively, tolerant to conductivity.

These West Virginia streams, like the reference streams used to develop WVSCI and GLIMPSS, were at one time thriving ecosystems, teeming with life that supported important functions for West Virginians and terrestrial and aquatic organisms alike. Downstream users rely on West Virginia's complex network of flowing streams for clean drinking water, fishing, recreation, and other important economic uses like tourism. These streams also serve cultural and spiritual purposes for West Virginians living near to and downstream from these once pristine rivulets. "Protecting these uses is the overriding purpose of West Virginia's water quality standards and the goal of the state's permit requirements." *Stillhouse*, 82 F. Supp. 3d at 699.

The Court accordingly **FINDS** that Plaintiffs have established, by a preponderance of the evidence, that Fola has violated its permits by discharging into Shanty Branch and Elick Hollow high levels ionic pollution as measured by conductivity, which have caused or materially contributed to a significant adverse impact to the chemical and biological components of the stream's aquatic ecosystem, in violation of the narrative water quality standards incorporated into those permits.

The Court also **FINDS** Plaintiffs have not established, by a preponderance of the evidence, that Fola caused or materially contributed to impairment in Leatherwood Creek by discharging ionic pollution into Shanty Branch and Elick Hollow.

Finally, Plaintiffs' motion to amend the Complaint is **DENIED**, ECF No. 81, and Fola's motion to supplement the record or for a new trial is also **DENIED**, ECF No. 78.

The Court will contact the parties to schedule the relief phase of this litigation. The Court **DIRECTS** the Clerk to send a copy of this Order to counsel of record and any unrepresented parties.

ENTER: May 26, 2017

ROBERT C. CHAMBERS, CHIEF JUDGE